



OPINIÃO

As estatísticas oficiais e a nova ciência de dados

As sociedades atuais exigem uma dramática alteração na produção estatística, que envolve todas as etapas do processo, desde a recolha de dados até à disseminação da informação.

26 de abril de 2017, 6:25

As sociedades atuais necessitam mais do que nunca de estatísticas oficiais fiáveis e oportunas para suportar a tomada de decisão aos mais variados níveis. Consequentemente, o desafio que se coloca hoje aos institutos de estatística é o de produzir informação cada vez mais abrangente, com maior nível de detalhe e de granularidade, mais rapidamente e com mais qualidade. É necessário compreender que estes produtores são atualmente apenas mais um *player* de um ecossistema competitivo de produção de informação estatística, que inclui os grandes gigantes dos dados digitais, mas muitos outros operadores, incluindo privados que procuram criar valor sobre dados provenientes das mais variadas fontes de natureza não oficial.

É neste ecossistema que os INEs terão de operar e compreender as necessidades dos velhos e novos utilizadores, que são cada vez mais variadas, mas sobretudo cada vez mais exigentes. Têm igualmente de compreender que nesta sociedade os métodos tradicionais que suportavam as operações estatísticas (largamente baseadas em inquéritos) apresentam limitações crescentes que se prendem com as cada vez maiores taxas de não resposta a inquéritos, com a necessidade de aliviar o esforço que é exigido a cidadãos e empresas ao participar nestas operações, com a falta de

~~Este site utiliza cookies. Ao navegar no site, está a consentir a sua utilização. Saiba mais sobre o uso de cookies.~~
flexibilidade das mesmas e, não menos importante, com os elevados custos que estas acarretam.



Estes desafios exigem uma dramática alteração na produção estatística, que envolve todas as etapas do processo, desde a recolha de dados até à disseminação da informação. A resposta passa inevitavelmente por: 1) uma maior utilização de dados administrativos (como sejam os dados da população, saúde, fiscais, da segurança social, da educação, do emprego e naturalmente os censitários); 2) o recurso às novas fontes digitais (o que vulgarmente se designa por *big data* e que inclui hoje dados provenientes de sistemas de telecomunicação, da atividade *web*, dos sensores a até de dados gerados pelos negócios; 3) e o recurso a novos modos de recolha de dados, nomeadamente à recolha de dados *online* em detrimento dos tradicionais modos baseados em entrevistas face-a-face. Esta revolução poderá contribuir para significativos benefícios com óbvias reduções no custo de recolha de dados, na redução dos tempos de disponibilização de informação estatística, no nível de detalhe das estatísticas produzidas e, não menos importante, na redução de externalidades negativas da tradicional produção estatística.

Claro que todas estas oportunidades exigem profundas mudanças metodológicas, que envolvem uma mudança de paradigma e de competências associadas ao tradicional estatístico, que será cada vez mais o do novo *data scientist*. Estas novas competências estão na interseção da tradicional estatística, mas também da computação, dos sistemas de informação e das ciências sociais. A utilização destas novas fontes exige cada vez mais a exploração de sistemas e tecnologias de informação e novos métodos de estimação, onde se destaca o enorme potencial da “estimação algorítmica” que tira partido de algoritmos computacionais (normalmente no âmbito da inteligência artificial) para trabalhar e integrar o grande volume de dados “em bruto” provenientes das referidas fontes administrativas e digitais. Esta mudança de paradigma implica também o domínio de novas técnicas de visualização que permitam aos utilizadores olhar para os dados de forma mais inteligível e intuitiva, mas sobretudo mais personalizada às suas necessidades. Seguramente que este caminho conduzirá rapidamente a uma nova era de estatísticas inteligentes capazes de se adaptar à dinâmica dos dados em tempo real e também ao perfil e necessidades dos utilizadores.

Só desta forma os tradicionais produtores de estatísticas oficiais poderão manter a sua relevância, e assumir o seu papel de principal agente dinamizador e credibilizador do ecossistema da informação estatística.

